

Time-varying lag cointegration

Philip Hans Franses*

Econometric Institute, Erasmus School of Economics, The Netherlands

ARTICLE INFO

Article history:

Received 1 July 2020

Received in revised form 28 October 2020

JEL classification:

C32

C51

Keywords:

Cointegration

Leads

Lags

Estimation

ABSTRACT

This paper proposes an alternative estimation method for cointegration, which allows for variation in the leads and lags in the cointegration relation. The method is more powerful than a standard method. Illustrations to annual inflation rates for Japan and the USA and to seasonal cointegration for quarterly consumption and income in Japan shows its ease of use and empirical merits.

© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper introduces an alternative estimation method for cointegration. This method makes use of the fact that if two integrated variables y_t and x_t are cointegrated, that then y_t and x_{t+1} and y_t and x_{t-1} are also cointegrated. This fact can be used to estimate an error correction variable $y_t - \theta x_{t-lag_t}$, where the variable lag_t can take values $-1, 0$, or 1 , throughout the sample $t = 1, 2, 3, \dots, T$.

Commonly applied methods to test for cointegration analyze the variables y_t and x_t at the same time, that is, these methods consider $y_t - \theta x_t$. There seems however no a priori reason to only allow for a contemporaneous cointegration relation. In fact, it may be restrictive to have shocks ε_t to hit the two variables always exactly at the same moment. Also, it may be that sometimes y_t leads and that sometimes x_t leads. A visual impression of this feature is presented in Fig. 1, which depicts annual CPI (Consumer Price Index) based inflation for Japan and the USA, for the sample 1960 to 2015. For some years, these two series seem to move together, like around 1975 and 1980, but in other periods, the USA inflation rates increase or decrease one year earlier than inflation in Japan does, while in other periods it is the other way around.¹

The outline of this paper is as follows. Section 2 shows that cointegration is not dependent on the specific lead and lag structure across the nonstationary variables. In fact, the same cointegration relation is found across various cases. Section 2 further provides a detailed illustration using annual CPI based inflation for Japan and the USA. Section 3 presents results of a few simulation experiments where it is examined what happens if the Data Generating Process (DGP) includes an error correction term like $y_{t-1} - \theta x_{t-1-lag_t}$, where lag_t can take values $-1, 0$, or 1 , and when the standard Engle and Granger [1] test method is used. It is found that the standard statistical method loses power. Section 4 addresses seasonal cointegration [2], where the new estimation method may be even more useful, because the seasonal random walk implies that “summer can become winter”. This feature makes it even more unlikely that exactly the same error process makes these changes to happen at the same time. Section 5 concludes with various avenues for further research.

* Correspondence to: Econometric Institute, Erasmus School of Economics, POB 1738, NL-3000 DR Rotterdam, The Netherlands.

E-mail address: franses@ese.eur.nl.

¹ Note that these two variables are non-stationary as the Dickey–Fuller test value for Japan is -2.567 and for USA it is -1.830 , which compared with the 5% critical value of -2.918 (only intercept, no trend) suggests that both variables are not stationary.

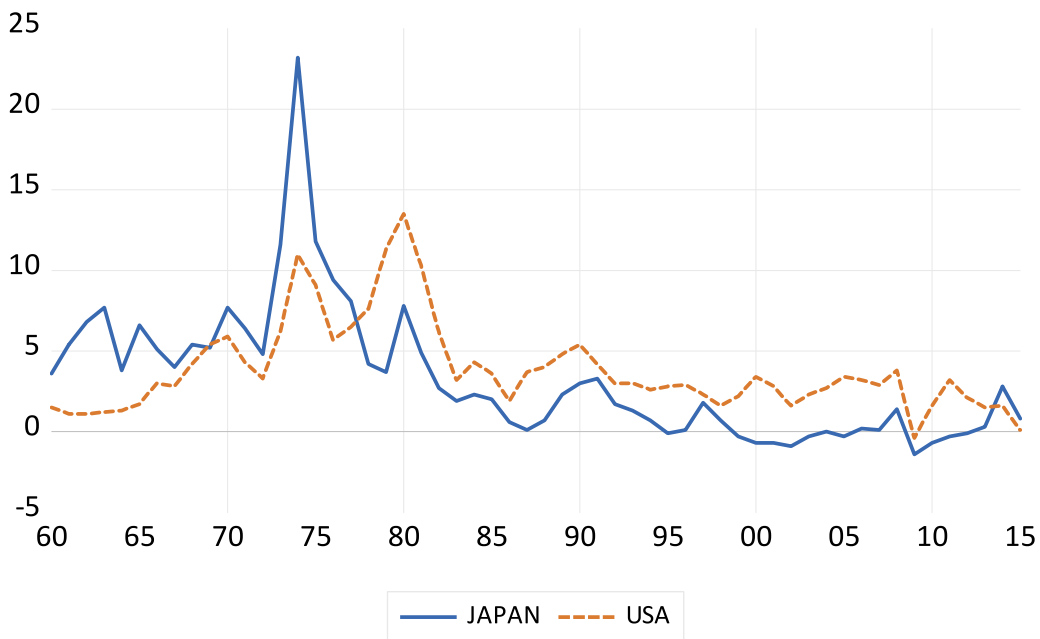


Fig. 1. Annual CPI based inflation rates for Japan and the USA.

2. Cointegration and lags

This section considers an exemplary case where two variables are cointegrated and can be captured by a single equation Error Correction Model (ECM).

Consider two variables y_t and x_t , for $t = 1, 2, 3, \dots, T$, and suppose that these two variables are connected via the Autoregressive Distributed Lag (ADL) model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \tag{1}$$

where ε_t is a standard zero mean white noise process with common variance σ_ε^2 . To save notation, an intercept is excluded.

2.1. Error correction model

It is assumed that the two variables are integrated of order 1, $I(1)$. That is, $y_t - y_{t-1}$ and $x_t - x_{t-1}$ are stationary, $I(0)$. It is further assumed that there is a single cointegration relationship between the two variables. In that case, ADL model in (1) can be written as an ECM like

$$y_t - y_{t-1} = (\alpha_1 + \alpha_2 - 1) \left(y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha_1 - \alpha_2} x_{t-1} \right) - \alpha_2 (y_{t-1} - y_{t-2}) + \beta_0 (x_t - x_{t-1}) + \varepsilon_t, \tag{2}$$

where the cointegration relation is

$$y_t - \frac{\beta_0 + \beta_1}{1 - \alpha_1 - \alpha_2} x_t$$

and where $(\alpha_1 + \alpha_2 - 1)$ in (2) is the adjustment parameter.

An alternative way of writing an ECM from the ADL model in (1) is

$$y_t - y_{t-1} = (\alpha_1 + \alpha_2 - 1) \left(y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha_1 - \alpha_2} x_t \right) - \alpha_2 (y_{t-1} - y_{t-2}) - \beta_1 (x_t - x_{t-1}) + \varepsilon_t \tag{3}$$

which shows that the parameter in the cointegration relation does not change when the long-run relation holds between y_t and x_{t+1} . A third way of writing the ADL model in (1) as an ECM is

$$y_t - y_{t-1} = (\alpha_1 + \alpha_2 - 1) \left(y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha_1 - \alpha_2} x_{t-2} \right) - \alpha_2 (y_{t-1} - y_{t-2}) + \beta_0 (x_t - x_{t-1}) + (\beta_0 + \beta_1) (x_{t-1} - x_{t-2}) + \varepsilon_t, \tag{4}$$

and this shows that the very same cointegration relation holds for y_t and x_{t-1} .

2.2. Two inflation series

One method to test for cointegration is the Engle and Granger [1] two-step method. The first step amounts to a regression of y_t on a constant and x_t (or x_{t-1} or x_{t+1}). The second step is to run a Dickey-Fuller (DF) unit root test on the residuals. One can also look at the Durbin Watson (CRDW) test statistic as recommended in Sargan and Bhargava [3].²

The first regression for the inflation series (for the effective sample 1961–2015 for comparison purposes) corresponding to (2) results in

$$Japan_t = -0.196 + 0.885USA_t$$

with a CRDW value of 0.449, and where the DF test (no lags of the first differences needed) obtains the value -2.775 . The 5% critical value for this test is -3.37 , and hence the null hypothesis of a unit root is not rejected. There seems to be no evidence of cointegration. The second regression corresponding with (3) results in

$$Japan_t = 0.306 + 0.769USA_{t+1}$$

with a CRDW value of 0.716, and a DF test value equal to -3.384 . This suggests the rejection of the unit root null hypothesis at the 5% level, and hence now there is evidence for cointegration. The third regression corresponding with (4) results in

$$Japan_t = 0.775 + 0.631USA_{t-1}$$

with a CRDW value equal to 0.621 and a DF test value of -3.185 , which now suggests the absence of cointegration.

Clearly, in this empirical case, the results give mixed evidence for cointegration.

2.3. An alternative estimation method

An alternative estimation method resorts to a time-varying lag cointegration model, that is, to consider the cointegration relation as

$$y_t - \theta x_{t-lag_t}$$

where $lag_t \in \{-1, 0, 1\}$. One way to determine the lag_t variable is to consider the three regressions above, compute for each regression the residuals, compare the absolute values of these residuals, and set the lag at each time t at the value that corresponds with the smallest absolute residual.

An application of this simple method to two inflation series, results in the lag_t variable, depicted in Fig. 2. The associated cointegration regression is

$$Japan_t = -0.915 + 1.155USA_{t-lag_t}$$

with a CRDW value equal to 0.804 and a DF test value of -3.739 , which provides strong evidence for the presence of cointegration amongst the two variables.³

A suitable error correction model turns out to be

$$Japan_t - Japan_{t-1} = -0.039 - 0.371 (Japan_{t-1} + 0.915 - 1.155USA_{t-1-lag_t}) + 0.903(USA_t - USA_{t-1})$$

with estimated standard errors 0.289, 0.107 and 0.171, respectively. Comparing the parameters with their standard errors suggests that the long-run and short-run relations between the inflation series are about equal to 1.

² The 5% critical value is 0.7.

³ Note that the Johansen [4] test gives a significant first eigenvalue and the cointegration coefficient is calculated as 1.250, which provides additional support for cointegration.

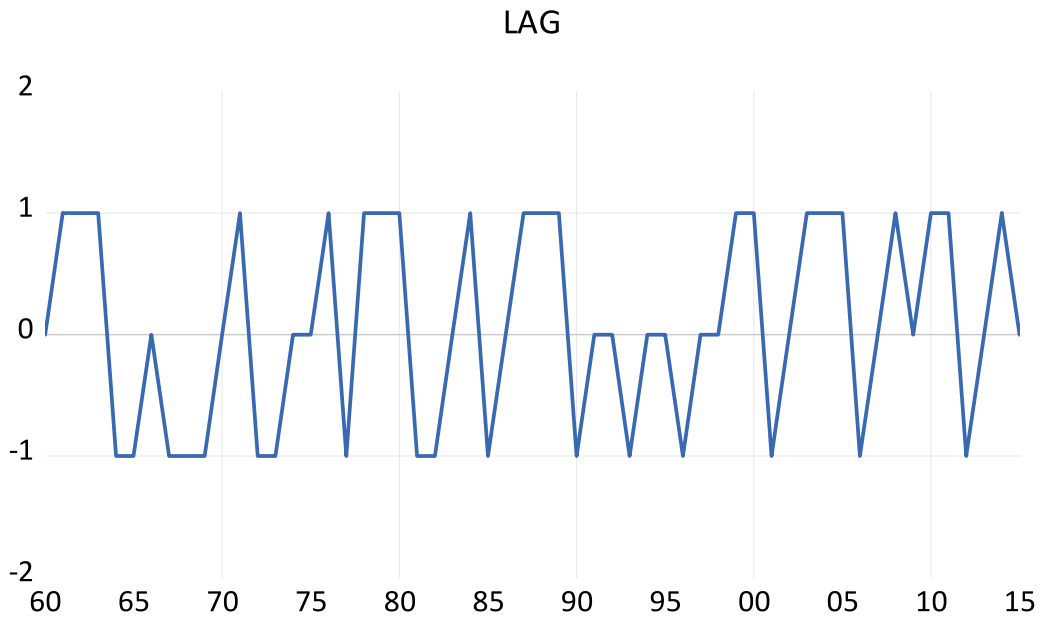


Fig. 2. Estimated lag structure in the time-varying lag cointegration regression for annual inflation series.

3. Simulation experiments

In this section a simulation experiment is considered to examine what happens if there is a time-varying lag cointegration structure, while one relies on the standard Engle–Granger single equation test. This latter test assumes that cointegration occurs at time t for both variables. Assume that the Data Generating Process (DGP) is

$$y_t - y_{t-1} = (\rho - 1)(y_{t-1} - x_{t-1-lag_t}) + u_t$$

$$x_t - x_{t-1} = w_t$$

with $u_t \sim N(0, 1)$ and $w_t \sim N(0, 1)$, $y_0 = 0$, $x_0 = 0$, $t = 1, 2, \dots, 120$, and where

$$lag_t \in \{-1, 0, 1\}$$

Per observation t , the value of the lag is drawn from a multinomial distribution with equal probabilities for the outcomes $-1, 0$ and 1 . Next, consider the regression

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

where the parameters are estimated using Ordinary least Squares (OLS). The estimated residuals are stored, and next the test regression

$$\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1} = \delta \hat{\varepsilon}_{t-1} + \vartheta_t$$

is considered. The t ratio for δ is computed and compared against the 5% critical value -3.37 .

Table 1 gives the number of cases (out of 10000 replications) that the t test value indicates rejection of the null hypothesis. The first column in Table 1 concerns the case where there is no variation in the lag, and this provides the statistical power of the standard Engle–Granger test method. The next column gives the rejection frequencies in case the DGP has a lag structure with $lag_t \in \{-1, 0, 1\}$, each with probability one-third. Obviously, there is a loss of power. One would expect the loss of power to increase when there is more variation in the lags, like for example $lag_t \in \{-1, 0, 1, 2\}$, each with probability one-fourth. And indeed, the final column of Table 1 indicates a further loss of power.

4. Seasonal cointegration

When analyzing quarterly nonstationary data, each of which obeys a seasonal random walk, that is, $y_t - y_{t-4}$ and $x_t - x_{t-4}$ are stationary, one may want to examine whether there is cointegration across $y_t + y_{t-1} + y_{t-2} + y_{t-3}$ and $x_t + x_{t-1} + x_{t-2} + x_{t-3}$, see Engle et al. [2] and Hylleberg et al. [5]. As a seasonal random walk implies that seasons may switch places, that is “summer can become winter”, it may even be more unlikely in practice that exactly

$$y_t + y_{t-1} + y_{t-2} + y_{t-3} - \theta(x_t + x_{t-1} + x_{t-2} + x_{t-3})$$

Table 1

The DGP is $y_t - y_{t-1} = (\rho - 1)(y_{t-1} - x_{t-1-lag_t}) + u_t$ and $x_t - x_{t-1} = w_t$ with $u_t \sim N(0, 1)$ and $w_t \sim N(0, 1)$, $y_0 = 0, x_0 = 0, t = 1, 2, \dots, 120$. The simulations are based on 10000 replications of the Engle-Granger two-step method.

ρ	$lag_t = 0$	$lag_t \in \{-1, 0, 1\}$	$lag_t \in \{-1, 0, 1, 2\}$
0.50	10000	10000	10000
0.55	10000	10000	9997
0.60	10000	9995	9989
0.65	9991	9978	9906
0.70	9913	9783	9512
0.75	9331	8941	8279
0.80	7430	6767	5925
0.85	4412	3975	3445
0.90	1817	1682	1388
0.95	670	659	617

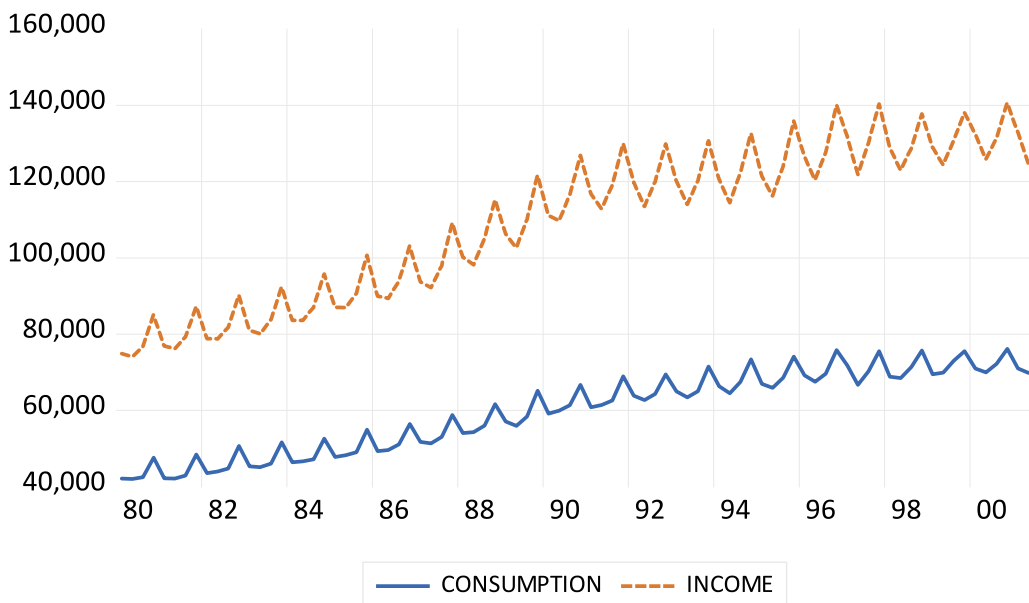


Fig. 3. Quarterly consumption and income in Japan, 1980Q1 to 2001Q2.

is stationary. An alternative cointegration relation can now be

$$y_t + y_{t-1} + y_{t-2} + y_{t-3} - \theta(x_{t-lag_t} + x_{t-1-lag_t} + x_{t-2-lag_t} + x_{t-3-lag_t})$$

where $lag_t \in \{-3, -2, -1, 0, 1, 2, 3\}$. The same method as above can now be considered, where various auxiliary regressions are run, and the size of the absolute residuals can be compared across the regression model, to select the lag variables.

4.1. Consumption and income in Japan

To illustrate the usefulness of time-varying lag cointegration, consider the quarterly data for consumption and income in Japan, for 1980Q1 to 2001Q2 in Fig. 3. A trend and a seasonal pattern are clearly visible for both series. The data are analyzed after taking natural logs, and y_t is the log of consumption and x_t is the log of income.

An Engle-Granger type regression assuming a contemporaneous cointegration relation results in

$$y_t + y_{t-1} + y_{t-2} + y_{t-3} = -0.937 + 0.968(x_t + x_{t-1} + x_{t-2} + x_{t-3})$$

The CRDW value is 0.060, which indicates the absence of cointegration. This is confirmed by an Augmented Dickey-Fuller test (one additional lag of the first differences) on the residuals of the test regression, which obtains the value -2.571 . This is not significant according to the critical values reported in Engle et al. [2].

When a time-varying lead and lag structure is allowed, the obtained time-varying lag structure is depicted in Fig. 4. This lag structure shows much variation. The alternating lag regression now gives

$$y_t + y_{t-1} + y_{t-2} + y_{t-3} = -0.240 + 0.953(x_{t-lag_t} + x_{t-1-lag_t} + x_{t-2-lag_t} + x_{t-3-lag_t})$$

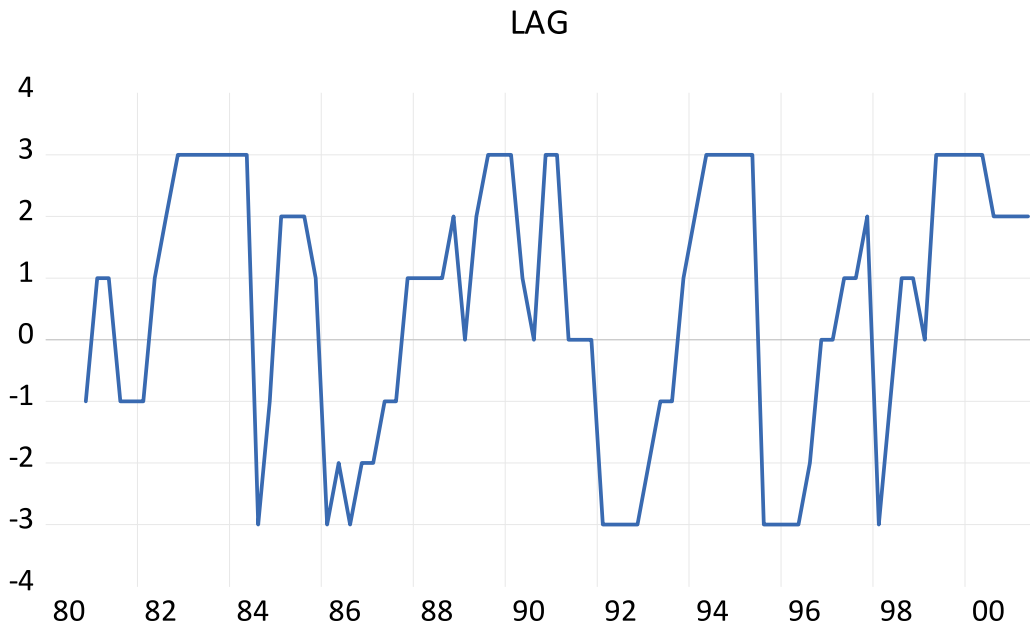


Fig. 4. Estimated lag structure in the time-varying lag seasonal cointegration regression.

with a CRDW value of 0.616, and a Augmented Dickey–Fuller test value of -3.815 . Hence, now there is evidence of cointegration.

A suitable error correction model⁴ for these two series turns out to be

$$y_t - y_{t-4} = 0.040 + 0.628(y_{t-1} - y_{t-5}) + 1.088(x_t - x_{t-4}) - 0.879(x_{t-1} - x_{t-5}) \\ - 0.043(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4} - 0.930)(x_{t-1-lag_t} + x_{t-2-lag_t} + x_{t-3-lag_t} + x_{t-4-lag_t})$$

where the parameters are estimated using Nonlinear Least Squares, and the associated standard errors are 0.065, 0.078, 0.081, 0.105, 0.013 and 0.033, respectively. The t test value on the adjustment parameter -0.043 is -3.443 , which is significant at the 5% level.

5. Conclusion

This paper has proposed an alternative estimation method for cointegration, which allows for variation in the leads and lags in the cointegration relation. The method is more powerful than a standard method. Illustrations to cointegration amongst annual inflation rates for Japan and the USA and to seasonal cointegration for quarterly consumption and income in Japan shows its ease of use and empirical merits. It is also demonstrated that the standard test method for cointegration does not find evidence of cointegration in both cases.

Various extensions of the new estimation method seem obvious. One could want to allow for more than two variables and consider the application of the Johansen [4] method. Further, one can consider nonlinear and time-varying cointegration. Also, one may want to allow for structural breaks and shifts in trend, which may not need to happen at the same moment in time for each series.

References

- [1] R.F. Engle, C.W.J. Granger, Co-integration and error correction: Representation, estimation, and testing, *Econometrica* 55 (1987) 251–276.
- [2] R.F. Engle, C.W.J. Granger, S. Hylleberg, H.S. Lee, Seasonal cointegration: The Japanese consumption function, *J. Econometrics* 55 (1993) 275–298.
- [3] J.D. Sargan, A. Bhargava, Testing residuals from least squares regression for being generated by the Gaussian random walk, *Econometrica* 51 (1983) 153–174.
- [4] S. Johansen, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59 (1991) 1551–1580.
- [5] S. Hylleberg, R.F. Engle, C.W.J. Granger, B.S. Yoo, Seasonal integration and cointegration, *J. Econometrics* 44 (1990) 215–238.
- [6] H.P. Boswijk, Testing for an unstable root in conditional and structural error correction models, *J. Econometrics* 63 (1994) 37–60.

⁴ This is a conditional error correction model, as proposed in Boswijk [6].